Parichay Maharaja Surajmal Institute Journal of Applied Research

Transforming Intelligence: A Comprehensive Review of Tiny Machine Learning (TinyML)

Mr. Sundeep Kumar[1]

Abstract

Tiny Machine Learning (TinyML) stands at the intersection of AI and embedded systems, epitomizing a paradigm shift in intelligent computing. In this review paper, we explores the profound impact of TinyML, delving into its hardware and software requirements, diverse applications, inherent benefits, existing constraints, and the transformative potential it holds for the future of smart technologies.

Introduction

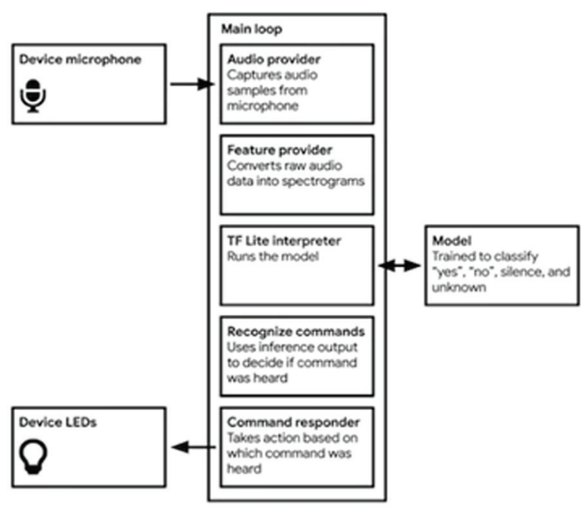
State-of-the-art deep learning AI systems usually demand extensive resources, such as large labeled datasets, significant computational power, and expertise from multiple AI specialists, for both training and inference processes (1). This poses a challenge for deploying these potent AI systems on edge devices. In the contemporary landscape of AI, the integration of ML algorithms with edge devices has become a pivotal focus. TinyML, is a revolutionary paradigm, transcending the boundaries of traditional computing. TinyML is represents an embedded ML technique that facilitates the deployment of ML applications on various affordable, resource- and power-limited devices (2).

This approach addresses minor issues that have the potential to create innovative opportunities across various technology fields. Its decentralized nature safeguards data privacy from the outset and operates efficiently on low power, enabling battery-powered usage. Moreover, it allows rapid testing and deployment. TinyML demands integrated solutions covering hardware, system, software, and application aspects, incorporating ML architectures, techniques, tools, and approaches capable of conducting on-device analytics at the cloud's extreme edge (3). TinyML can be utilized in energy-efficient systems like sensors or microcontrollers to execute automated functions (3).

Major Applications of TinyML

In the modern era, TinyML is part of our daily life, either we realize it or not. There are more than 3 billion small devices that are based on ML models (1). TinyML involves processing time-series data directly from sensors in real-time. TinyML techniques have found successful applications in various commercial products, deployed in domestic, office, and industrial settings (4, 5).

1. **Keyword Spotting:** There is an audio wake-word detection model implemented in Google, iOS and Android devices which turn on these device upon hearing these words, 'OK Google', 'Hey Siri', 'Alexa' (4).



Components for a wake-word application. Image used courtesy of Zhitong Yan and Zhuowei Han

2. **Visual Wake Word:** This technique expands on keyword spotting for images and has been adopted in various applications available in the market that utilize the visual wake word technique. Example Google Lens. Other common applications include camera sensors capable of detecting people in a room. Such systems can automatically adjust lighting by turning it off when no person is detected, or they can be employed for security purposes (5).

3. **Anomaly Detection:** This technique is utilized to identify unexpected events. The anomaly detection technique is based on an unsupervised learning approach, wherein the model must identify patterns within unlabeled data. This technique is primarily used in industries to detect malfunctioning on factory machines (5).

[1] *Assistant Professor, Department of Computer Application, Maharaja Surajmal Institute (affiliated to GGSIP University, Delhi), C-4, Janakpuri, New Delhi-58.*

Key Industries Served by TinyML

TinyML applications extend across various sectors, particularly those reliant on IoT networks and data [6]. Common applications of TinyML include computer vision, visual wake words, keyword spotters, predictive maintenance, gesture recognition, and industrial machine maintenance, among others, catering various markets/ industries.

Agriculture: TinyML devices enable the real-time monitoring and collection of agricultural and livestock data. TinyML applications in agriculture involve analyzing data from sensors and drones to monitor soil quality, crop health, and pest infestations. Imagimob, a Swedish edge AI product company, has developed a ML platform for edge devices. More than 50 organizations across the Europe have partnered with Imagimob to explore how TinyML technique can facilitate effective management of crops and livestock [6].

Predictive Maintenance in Manufacturing: TinyML enables predictive maintenance solutions in manufacturing equipment. By analyzing sensor data locally, it predicts machinery failures before they occur, minimizing downtime and optimizing maintenance schedules. TinyML can be implemented on low-powered devices to conduct continuous monitoring of machines, identifying malfunctions and predicting issues before they occur. Example is Ping Service, developed a monitoring device designed to consistently analyze the acoustic patterns of wind turbine blades, alerting users to any alterations or damage detected [6].

Customer Experience: To offer personalized services to their customers, companies need to understand better their customers' behavior to target them with personalized ads and messages. TinyML applications empower enterprises to understand user contexts, encompassing their behaviors and interactions [6].

Environmental Monitoring and Conservation: TinyML equipped sensors are deployed in environmental monitoring systems to track air quality, detect forest fires, and monitor wildlife behavior. This data aids conservation efforts and facilitates rapid response to environmental threats [7].

Gesture Recognition in Human-Computer Interaction: TinyML powers gesture recognition interfaces in applications such as gaming consoles and virtual reality systems. By interpreting hand gestures, it enhances user experience, making interactions intuitive and immersive [8].

Tools Required for TinyML

TinyML's unmatched prowess lies in its nuanced interplay between specialized hardware and optimized software architectures. A meticulous dissection of the hardware requirements reveals the intricacies of microcontrollers, their limited computational resources, and the integration of purpose-built accelerators.

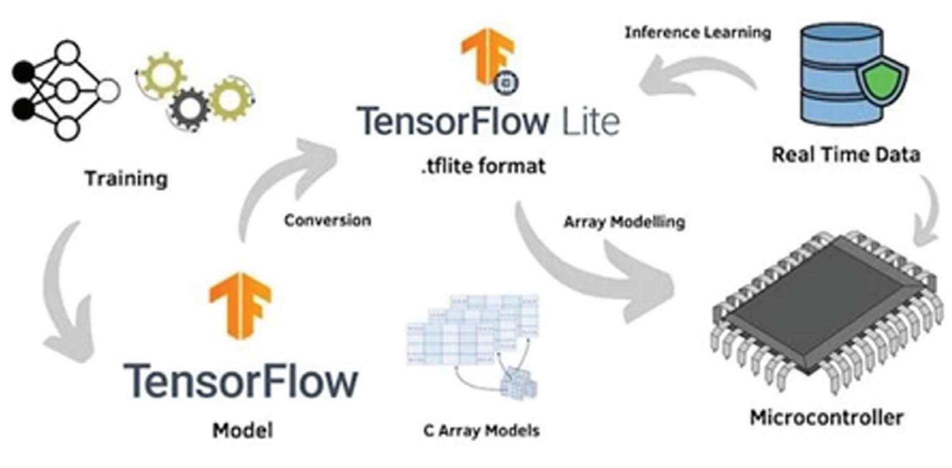
Simultaneously, the software landscape unfolds, exploring the nuances of model quantization, and compression algorithms, orchestrating a symphony of efficiency that allows sophisticated ML models to operate within the confines of miniature devices.

Microprocessors: TinyML devices are equipped with MCUs and digital signal processors (DSPs) to meet 1 mW objectives. The devices are typically Cortex-M based, featuring limited RAM, comparable flash storage, and clock rates ranging in the tens of MHz [9].

Batteries: According to Pete Warden, widely accepted founder of TinyML, the objective for TinyML should be achieving an energy usage level below 1 mW. This particular value is chosen because a consumption of 1 mW enables a device to operate for several months to a year on a standard coin battery. There are a few options such as small Li-Po batteries, coin batteries, and energy harvesting devices when selecting power sources for TinyML applications [9].

Python: Python is preferred language, having various libraries and frameworks, to build ML models, for example TensorFlow, PyTorch, Keras. However C, C ++, or Java to build ML models can be utilized for the same [5, 9].

TensorFlow: This is an open source ML framework developed by Google, having a Python front-end with optimized C++ code at the core. It provides pre-trained ML models designed for everyday use. With Tensorflow Lite, ML models can be created without connecting to the Internet [4, 8]. TF Lite Micro is utilized for data acquisition, preprocessing, model architecture, training, evaluation, optimization and quantization [4].



The TensorFlow Lite Micro workflow. Image used courtesy of Saumitra Jagdale

Apart from theses, sensors, (camera, microphone, etc.) and Bluetooth Low Energy (BLE) connectivity components are also available in TinyML devices [9].

Advantages of TinyML

TinyML offers a multitude of advantages, reshaping the landscape of intelligent computing:

Cost effective: TinyML devices operate on affordable microcontrollers, utilizing small batteries and ensuring minimal power consumption [4].

Real-time Processing: Empowering devices with instantaneous decision-making capabilities, fostering applications where split-second responses are quintessential.

Energy Efficiency: By minimizing data transmission to external servers, TinyML devices conserves energy, extending device battery life and reducing overall power consumption in various applications. Solutions based on MCUs consume significantly lower power even under heavy workloads [9].

Data Privacy and Security: Local data processing enhances privacy by reducing reliance on cloud-based services, thereby mitigating potential security risks associated with data transmission over networks [4].

Key Challenges

While TinyML holds immense promise, it is not without challenges.

Heterogeneity in Tools: The TinyML domain, akin to the IoT sector, encompasses diverse hardware components and algorithms, posing challenges in understanding the tradeoffs between various TinyML implementations [9].

Battery Power Consumption: Under optimal circumstances, a 2Ah battery capacity should ideally sustain a lifespan exceeding ten years, assuming a power consumption rate below 12µA. However, in real-world scenarios, the cumulative current consumption rises to approximately 176.4 mA, significantly diminishing the battery's longevity[2].

Conclusion

In recent years, TinyML has garnered momentum in diverse industries, thanks to the evolution of supportive hardware and software ecosystems and also has more potential to unlock an entire new realm of smart applications in industrial and consumer sectors [4, 9]. Its ability to process complex algorithms on devices with minimal resources opens avenues for innovation in healthcare, industrial automation, smart homes, and beyond. TinyML stands as a testament to human ingenuity, unraveling the boundless possibilities of intelligent computing in a miniature world.

References

1. https://www.thoughtworks.com/en-in/insights/blog/machine-learning-and-ai/tinyml-the-next-big-thing
2. TinyML: Tools, Applications, Challenges, and Future Research Directions, Mar 23, 2023, by Rakhee Kallimani, Krishna Pai, Prasoon Raghuwanshi, Sridhar Iyer and Onel L. A. López
3. https://medium.com/tech-cult-heartbeat/what-is-tinyml-and-why-does-it-matter-f5b164766876
4. https://www.allaboutcircuits.com/technical-articles/what-is-tinyml/
5. TinyML: From Basic to Advanced Applications, June 21, 2021, by Marc Monfort Grau
6. https://www.datacamp.com/blog/what-is-tinyml-tiny-machine-learning
7. How TinyML Can be Leveraged to Solve Environmental Problems: A Survey, Nov 2022, by Hatim Bamoumen, Anas Temouden, Nabil Benamar, Yousra Chtouki
8. https://www.hackster.io/leonardocavagnis/gesture-recognition-tinyml-for-8-bit-microcontroller-1cb0a8
9. https://www.artiba.org/blog/tinyml-the-future-of-machine-learning